P.D. 1997 (6)
p. 168-173

TECHNIQUES

Proteome analysis: from protein characterization to biological function

Stephen R. Pennington, Marc R. Wilkins, Denis F. Hochstrasser and Michael J. Dunn

The sequencing of the genomes of several organisms, including human, is proceeding at a dramatic pace. However, the function of the protein products of many of the newly identified genes is unknown. Cells typically express several thousand different proteins, and each of these may be subject to a wide variety of dynamic posttranslational modifications, effectively increasing the number of different proteins present. Although not yet being undertaken on the same scale as the DNA-sequencing programmes, a number of projects aimed at characterizing all the PROTeins expressed by the genOME of an organism – the 'proteome' – are now under way. Here, the authors discuss some important recent developments in the comprehensive analysis of protein expression – proteome analysis – which together have put this approach at a watershed. Current progress suggests that the data it will generate may alter fundamentally our understanding of the molecular composition and function of cells.

Stephen Pennington is in the Dept of Human Anatomy and Cell Biology, New Medical School, Ashton Street, University of Liverpool, Liverpool, UK L69 3GE; Marc Wilkins and Denis Hochstrasser are in the Central Clinical Chemistry Laboratory, University Hospital of Geneva and the Medical Biochemistry Dept, University of Geneva, CH 1211 Geneva, Switzerland; and Michael Dunn is at the National Heart and Lung Institute, Imperial College School of Medicine, Heart Science Centre, Harefield Hospital, Harefield, Middlesex, UK UB9 6JH. E-mail: srpenn@liv.ac.uk

There is widespread awareness of the genome-sequencing projects that have been completed or are currently under way (Table 1). The recent announcement of the complete sequence of an archaeon (*Methanococcus jannaschii*)[1] means that comprehensive genome-sequence comparisons of bacteria, archaea and eukarya are now possible. The most ambitious of the as-yet-incomplete sequencing projects, the human genome project, was initiated in the USA in 1990 and now involves laboratories in the UK, France, Germany and Japan. It is estimated that it will cost in excess of $3 billion over a period of 15 years. Initially, genetic and physical mapping was a priority, but now the second phase – sequencing – has begun and has an anticipated completion date of a point during the first decade of the next millennium[2]. For obvious reasons, the mapping and sequencing of the human genome attracts considerable interest not least from commercial organizations, which see enormous (financial) potential in the information that the project will generate.

The progress of existing sequencing projects has been dramatic: they are producing data for new genes at an exciting and ever-increasing rate. Exciting because, for example, the recently accomplished complete sequence (~12.5 Mb) of the eukaryotic budding yeast *Saccharomyces cerevisiae* has revealed that approximately half of the genes (2964) are previously unidentified, and of these about 2300 show no significant sequence homology to known genes and hence are of completely unknown function[3]. Moreover, the function of many of the new genes that show some sequence homology to known genes remains elusive. It has become apparent, therefore, that this huge and rapidly growing repository of information represents a starting point for the more challenging aim of understanding completely how a simple eukaryotic cell such as *S. cerevisiae* functions[3,4]. This will undoubtedly be superseded by the quest for a similarly detailed understanding of how higher eukaryotic cells and a complex organism, the human, function. Attention will turn increasingly to the protein products of the genes.

Although less publicized than their genome counterparts, a number of proteome projects have been started, and in some cases significant progress has been made (see Table 1). In these projects, complex mixtures of proteins are resolved by two-dimensional (2-D) gel electrophoresis, and individual proteins characterized by a growing array of methods. Recent developments have made important improvements in both the speed and sensitivity with which such analyses can be undertaken (see Refs 5–8). In many instances, the methods used to characterize proteins present on a 2-D gel pattern are dependent on the existence of DNA sequence data – genome and proteome analysis are therefore (intimately) linked. The continued integration of proteome and genome data will have a fundamental impact on our understanding of the normal functioning of cells and organisms.

Reducing redundancy – the need to analyse protein expression

DNA sequence data alone cannot be used to build a 'molecular definition' of a cell. The DNA sequence data themselves reveal little or nothing about the level of expression of proteins, the protein isoforms that may be produced from each gene (by alternative splicing) or the extent to which protein(s) are posttranslationally modified. Nor can they provide information on the cellular or subcellular distribution of proteins. It is precisely this type of information that proteome analysis will provide.

One of the immediate impacts of genome sequence data is the ability to undertake mass screening of mRNA expression, and to this end a number of impressive approaches are being developed. Current progress includes the application of cDNA microarrays[9] and serial analysis of gene expression[10]. These should provide information on which mRNAs are expressed and therefore which proteins are likely to be expressed. If sufficient selectivity can be achieved, and recent reports suggest that it can[11], they should also provide important information on the presence of mutated forms of genes and so could be particularly useful for diagnostic screening. Strategies are also being developed to establish the biological function of unknown genes[12]. However, it is clear that even with these approaches it is necessary to undertake comprehensive analysis of protein expression. Fortunately, the past two or three years have seen important developments in a number of techniques now enabling this to be undertaken.

There are two (concurrent) phases to proteome analysis. First, all the proteins expressed by a particular organism, tissue or cell under 'normal' conditions are identified and their position on 2-D gels mapped – this defines the 'constitutive' proteome. The information can be used to generate a 2-D reference map and database for the organism, tissue or cell examined. A number of 2-D reference maps and databases (see Fig. 1) are being developed and are accessible through the World Wide Web and Internet (Table 2). In some cases, these databases are already impressive; thus, for example, of the estimated 3500 or so proteins expressed in human keratinocytes, approximately 30% (1082) have been identified and are recorded in the 2-D reference database for this cell[13]. Similarly, 630 protein spots present on 2-D gels of human plasma

 0962-8924/97/$17.00
PII: S0962-8924(97)01033-7

TABLE 1 - PROGRESS OF SELECTED GENOME AND PROTEOME PROJECTS[a]

Organism	Genome size (Mb)	Estimated proteome size (gene number)	Protein entries in SWISS-PROT	Proteins annotated on 2-D maps
Mycoplasma genitalium[b]	0.58	470	344	>50
Haemophilus influenzae[b]	1.83	1743	1578	>100
Escherichia coli[b]	4.72	4000	3654	>300
Methanococcus jannaschii[b]	1.66	1738	62	-
Saccharomyces cerevisiae[b]	12.5	6000	3720	>250
Dictyostelium discoideum	70	12 500	217	>50
Arabidopsis thaliana	70	14 000	541	50
Caenorhabditis elegans	80	17 800	1030	-
Homo sapiens	2900	60 000–80 000	3719	>1500

[a]Estimated genome size, estimated proteome size, number of protein sequences in SWISS-PROT (Release 33, July 1996) and approximate number of proteins of known identity on two-dimensional reference maps for some model organisms for which genome-sequencing projects are under way or completed.
[b]Genome sequences that have been completed. Note that there is a lag time for submitted sequences to appear in SWISS-PROT as all submissions are checked thoroughly and annotated prior to entry in the database. It is noteworthy that posttranslational modifications will increase the total number of individual proteins expressed from a genome; for bacteria, there may be a 1.2- to 1.3-fold increase over gene number, perhaps 3-fold for yeast and as much as 10-fold for human.

have been identified (Table 2, SWISS-2DPAGE). The achievements of these projects support the feasibility of proteome analysis.

The 2-D reference maps will serve as a foundation on which the changes in pr teome expression will be analysed. In the longer term, it is anticipated that all this information will be readily accessed and integrated with the other databases. More tangibly, the 2-D reference maps and databases should enable other groups to identify proteins on 2-D gels by computer-driven comparison of an image of their 2-D gels with those in the databases (see Fig. 1). For example, those studying t xic responses of the liver to drugs[14] may exploit a liver 2-D reference map (Table 2) to identify the proteins that change in response to individual treatments. To some extent, this may ultimately avoid unnecessary duplication of the effort required to characterize individual protein spots on 2-D gels.

In its 'second phase', proteome analysis involves the measurement of changes in the proteome that occur in response to changing physiological conditions. This includes analysis of the quantitative changes in protein expression, changes in posttranslational modifications and, where appropriate, chang s in subcellular localization of proteins. At least initially, this is likely to be focused on groups or clusters of proteins rather than the entire proteome. It may be that much f this analysis will rely on existing antibody-based methods and therefore not be dependent on 2-D separati n of proteins.

Using 2-D electrophoresis (2-DE) to analyse protein expression – the power and problems of 2-DE

Much of proteome analysis is dependent on the ability to separate complex protein mixtures rapidly, and 2-DE is undoubtedly the most powerful method currently available to achieve this. Estimates of the number of polypeptides that can be resolved on a single gel vary between 2500–10 000 (see Ref. 8). Moreover, 2-DE, combined with appropriate computer analysis of gel images[15], enables quantitative analysis of the amount of individual proteins within such mixtures. The method also potentially allows simultaneous purification of thousands of proteins, unlike conventional protein-purification strategies, which are focused on the production of one or a few individual proteins. Moreover, the 2-DE-based approach allows proteins to be purified without any prior knowledge of their properties.

One of the biggest differences between conventional and 2-D protein purification is the amount of protein that is purified and therefore available for analysis. In the past, the low loading capacity of 2-D gels has severely limited its application as a purification step. Even with this limitation, significant characterization of proteins from 2-D gels has been achieved with methods such as immunoblotting, N-terminal and internal peptide sequencing, comigrati n f unknown pr teins with known proteins, or overexpression of homologous genes of interest in the organism under study (see Ref. 13). However, it is now possible to apply milligram quantities of protein mixtures onto immobilized pH gradient (IPG) isoelectric-focusing (IEF) gels, so producing micropreparative 2-DE and enabling the purification of microgram amounts of 200–300 proteins[15]. The more recent use of narrow-range IPG gels (1.0 pH units) allows up to 15 mg of protein to be applied to a single gel; thus, using a series of these gels (of different pH ranges) could allow the purification of as many as 3000 proteins at 5 μg per protein[16]. These improvements have extended the applicability of 2-DE for the preparation of proteins for analysis. However, it should be noted that some proteins are not readily solubilized by existing IEF buffers and that some high (>200 kDa) and low (<10 kDa) molecular mass proteins are not adequately resolved by 2-DE. Despite these limitations, it seems evident that 2-DE can be used to undertake comprehensive protein analysis, and recent developments have accelerated efforts to produce new improved methods.

Clearly, the potential success of identifying proteins by comparison of their positions on a 2-D gel with locations on a 2-D reference map rests on the consistency and quality of the 2-D separations. In the past, this has posed problems, but current protocols based on IPGs provide spatially and quantitatively reproducible 2-D patterns both within a laboratory and, more importantly, betwe n laboratories[17,18]. Despite these significant improvements, the r utine production f high-quality gels on the scale required for comprehensive documentation of proteome changes still requires extensive practical expertise. This potentially

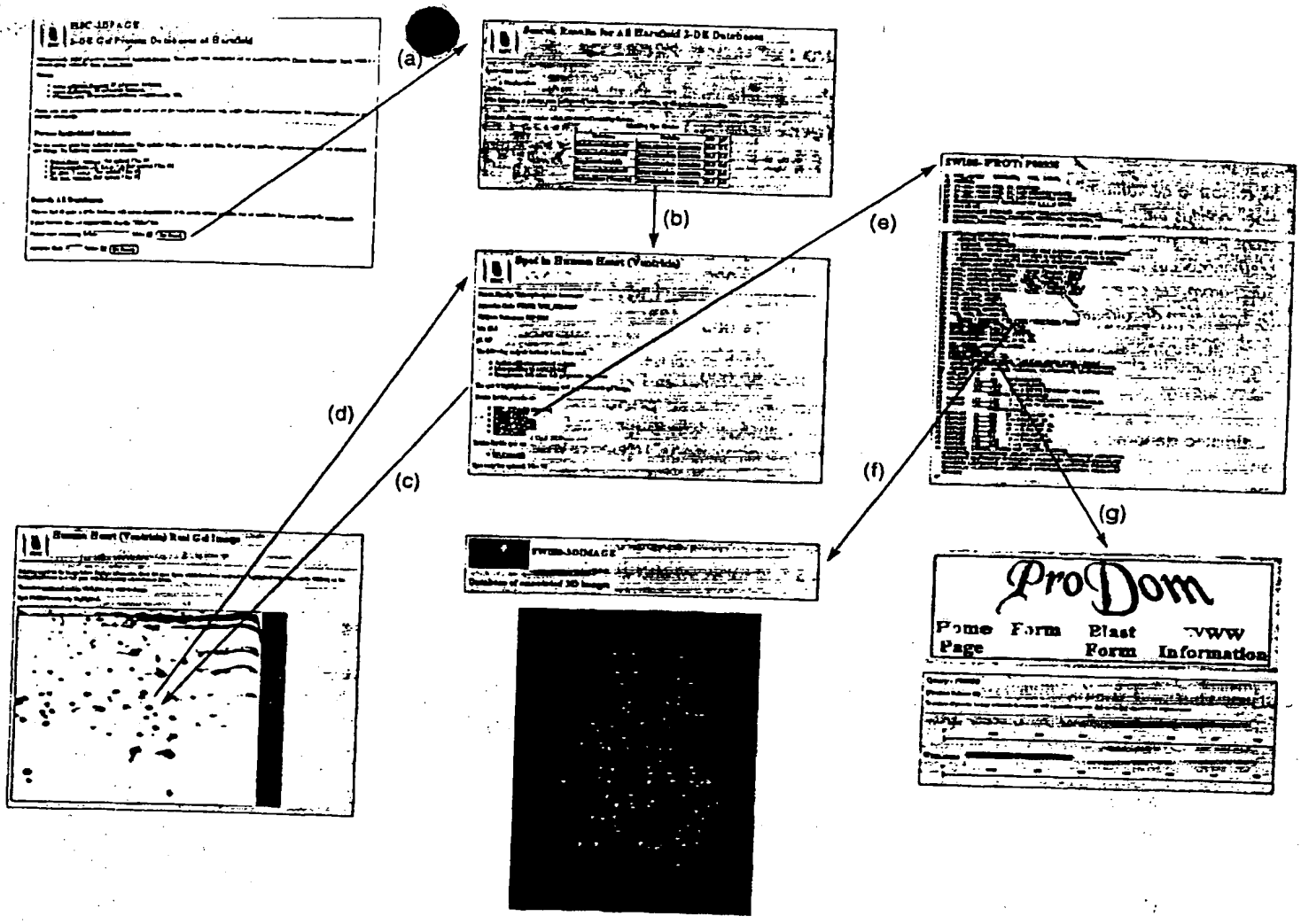

FIGURE 1

Some features of a two-dimensional database – the HSC-2DPAGE database of proteins of the cardiovascular system maintained in the Heart Science Centre, NHLI, Imperial College School of Medicine at Harefield Hospital, UK (see Table 2). (a) A full-text search for the term 'triose phosphate isomerase' (or 'triose') results in a list of the databases containing an entry for this term. (b) Selection of an item in the results table produces a display of the entry for the protein spot found in the human ventricle. (c) This is linked to the protein spot in the reference two-dimensional (2-D) gel image. (d) Database entries can be accessed directly by 'clicking' on a protein spot in the reference 2-D gel image. (e) The database has hyperlinks to other databases, and SWISS-PROT is used as the main index. (f,g) There are extensive links from SWISS-PROT to other databases, including GenBank, EMBL, SWISS-3DIMAGE (f), the Protein Domain Database (g), Medline, SWISS-2DPAGE, OMIM and PROSITE.

limits the applicability of 2-DE. A fully automated 2-D 'machine' would be very useful, and the development of such instrumentation is under way in several laboratories (see, for example, Ref. 19). The now obvious potential of proteome analysis will undoubtedly result in others attempting to develop automated instrumentation. Alternatively, the use of precast gels with both first-dimension IPG and second-dimension SDS–polyacrylamide gels in an all-in-one unit, requiring just a swapping of electrode positions to run the two dimensions, might provide highly reproducible separations that would be technically less demanding and perhaps more readily applicable to automation.

Recent breakthroughs in protein identification and characterization

Unequivocal identification of a protein is most convincingly achieved by direct determination of its primary sequence, and until recently protein identification relied almost entirely on Edman degradation. With the increase in DNA and protein sequence information (Table 1), now only limited N-terminal or internal sequence, typically 10–20 amino acids, may be sufficient to identify many proteins. Even so, sequencing by Edman degradation is limited by a relatively low sample throughput of usually one or two peptides per day. Increasingly, groups are now using identification strategies that involve the combination of many pieces of quickly, economically and more easily obtained data to identify proteins with confidence (see Refs 5–8 and 15). This has led to the development of integrated approaches using combinations of amino acid composition, peptide mass fingerprinting, Edman sequencing and tandem mass spectrometry (Table 3). These strategies also often use the estimated molecular weight and pI of the protein. So, for example, data for the amino acid composition or the masses (rapidly measured by mass spectrometry) of peptides generated by enzymatic cleavage of the protein can be combined with the molecular mass and pI of a protein spot to 'search' the SWISS-PROT

TABLE 2 – 2-D GEL DATABASES AND OTHER DATABASES OF RELEVANCE TO PROTEOME ANALYSIS[a]

Database	Description	Address
2-D gel databases		
Biobase	Keratinocytes, bladder cancer, etc. (Danish Centre for G nome Research)	http://biobase.dk/cgi-bin/celis
ECO2DBASE	*E. coli* (in NCBI repository)	ftp://ncbi.nlm.nih.gov/repository/ECO2DBASE
Heart-2DPAGE	Human myocardial (Berlin, Germany)	http://www.chemie.fu-berlin.de/user/pleiss/
HSC-2DPAGE	Human heart (Heart Science Centre, Harefield, UK)	http://www.harefield.nthames.nhs.uk/nhli/protein/index.html.
LSB	Human, mouse and rat liver, etc. (Large Scale Biology, USA)	http://lsbc.com/2dmaps/patterns.html
QUEST	RefS2 cells, mouse embryo, yeast (Cold Spring Harbor Lab., USA)	http://siva.cshl.org/
SWISS-2DPAGE	10 human maps including: liver, plasma, erythrocytes, platelets. Also, yeast, *E. coli* (University Hospital, Geneva, Switzerland)	http://expasy.hcuge.ch/ch2d/ch2d-top.html
Yeast	*S. cerevisiae*, *S. pombe*, etc. (Göteborg University, Sweden)	http://yeast-2dpage.gmm.gu.se
Yeast	*S. cerevisiae* (Proteome Inc., USA)	http://www.proteome.com/YPDhome.html
Protein sequence databases		
PDD	Protein Disease Database	http://www-pdd.ncifcrf.gov/
PIR	Protein Identification Resource – protein sequences	http://www.gdb.org/Dan/proteins/pir.html
ProDom	Protein Domains and families	http://www.sanger.ac.uk/~esr/prodom.html
PROSITE	Dictionary of Protein Sites and patterns	http://expasy.hcuge.ch/sprot/prosite.html
SWISS-PROT	SWISS-PROT annotated protein sequences	http://expasy.hcuge.ch/sprot/sprot-top.html
YPD	Yeast Protein Database	http://www.proteome.com
3-D protein structure databases		
PDB	Brookhaven Protein Data Bank	http://www.pdb.bnl.gov/
SCOP	Structural Classification Of Proteins	http://scop.mrc-lmbcam.ac.uk/scop
SWISS-3DIMAGE	3-D images of proteins (and other molecules)	http://expasy.hcuge.ch/sw3d/sw3d-top.html
Gen me and DNA databases		
EMBL	EMBL-EBI nucleotide sequences	http://www.ebi.ac.uk
GDB	Genome database	http://gdbwww.gdb.org/
GenBank	GenBank nucleotide sequences	http://gdbwww.gdb.org
GSDB	Genomic sequences	http:/www.ncgr.org/gsdb/gsdb.html
Useful indexes and lists of links to protein-related sites		
Amos's links	Amos Bairoch's WWW links page	http://www/amos_www_links.html
Bio-wURLd	Searchable collection of molecular biology URLs	http://ebi.ac.uk/htbin/wurld.pl
EBI	EBI homepage	http://www.ebi.ac.uk
ENTREZ		http://www3.ncbi.nlm.nih.gov/Entrez
ExPASy	Expert Protein Analysis System	http://expasy.hcuge.ch
Pedro's homepage	BioMolecular Research Tools	http://www.public.iastate.edu/~pedro/research_tools.html
SIR	Structural Information Resources: list of URLs on macromolecular structure	http://www.ucmb.ulb.ac.be/StrctResources

[a]Abbreviations: 2-D, two-dimensional; 3-D, three-dimensional; *E. coli*, *Escherichia coli*; *S. cerevisiae*, *Saccharomyces cerevisiae*; *S. pombe*, *Schizosaccharomyces pombe*; URL, uniform resource locator; WWW, World Wide Web.

database and generate a list of best-matching proteins. The probability of achieving a correct identification is increased dramatically by the combination of two or more approaches to generate the data used to search the sequence databases (Table 3). It might be assumed that protein identification by homology searching would not be applicable to organisms for which there is little DNA (or protein) sequence information; however, recent attempts at

TABLE 3 – INTEGRATED ANALYSIS FOR MASS SCREENING OF TWO-DIMENSIONAL-GEL-SEPARATED PROTEINS*

Potential order of analysis	Identification technique	Refs
1	Amino acid composition	22,30,31
2	Amino acid composition and short N-terminal Edman microsequencing	23
3	Peptide mass fingerprinting	31–33
4	Combination of amino acid composition and peptide mass fingerprinting	20,21
5	Mass spectrometry sequence tag	24
6	Extensive N-terminal Edman microsequencing	34
7	Internal peptide Edman microsequencing	35
8	Microsequencing by mass spectrometry (electrospray ionization, post-source decay MALDI-TOF)	25,26

*Rapid and inexpensive techniques are used as a first step in protein identification, and slower more expensive techniques then used, if required. The potential order of analysis will change with advances in technology (particularly in mass spectrometry).

cross-species mapping have proved successful, albeit on a limited scale[20]. The results of these integrated approaches, which can be undertaken with as little as 250 ng of protein, are dramatic both in terms of the success rate of positive identifications and the potential throughput of samples (see Ref. 22). The methods also offer the potential to search the expressed-sequence tag (EST) databases such that a limited amount of rapidly obtained data (i.e. sequence tag) from a protein spot of interest on a 2-D gel could be quickly 'converted' into a large amount of sequence information for the corresponding (previously uncharacterized) gene. However, while rapid and relatively straightforward, these approaches do not always provide unambiguous identification of a protein and are of course dependent on appropriate sequences being present in sequence databases. There remains, therefore, an important place for extensive protein sequencing by Edman degradation or more recently developed mass-spectrometry techniques. The most exciting of these has been the use of nano-electrospray tandem mass spectrometry to sequence a 2-D-separated and silver-stained protein (see article in this issue by A. Lamond and M. Mann, and Ref. 26). It is likely that the sensitivity and speed of protein characterization from 2-D gels will increase further and that the cost of each protein spot characterization will decline. Clearly, this will be necessary if the products of the 60 000–80 000 human genes, including all posttranslationally modified forms, are to be included in protein databases in the foreseeable future.

Integration of protein sequence data with genome sequence databases

The effective use of DNA sequence information is facilitated by well-maintained and readily accessible databases; the same will be true of proteome data. A number of 2-D databases have been established (see Table 2). In these databases, the protein spots on images of 2-D gels can be used to identify the protein (if its identity is known) and to provide a direct link to its protein and DNA sequence data within the sequence databases and to structural information (see Fig. 1). Access to the 2-D databases may also be achieved from the sequence databases. As 2-D databases grow in numbers of entries, so too will the application of the information and their integration with other databases, including those that document protein changes in disease states (Table 2).

Applications of proteome analysis

The potential applications of proteome analysis are numerous. At its most straightforward, proteome data will be used to confirm open-reading frames in DNA sequences. It may also provide information on possible alternatively spliced forms of mRNAs. More importantly, by comprehensively monitoring changes in protein abundance, it may be possible to assign a function to novel gene products by measuring changes in the proteome after overexpression or knockout of the gene of interest. Even if the function of a gene is known, or suspected, analysis of the effect its deletion on tissue/cell protein expression will prove valuable, particularly for the analysis of transgenics where redundancy in protein function may result in little apparent effect on phenotype. Alternatively, analysis of changes in protein expression that accompany large changes in phenotype of the transgenic may be very revealing.

More obvious are the applications of proteome analysis to measurement of changes in protein expression during development, through the cell cycle, during apoptosis, in diseases and in response to extracellular stimuli (hormones, cytokines, stress), drug treatment and the application of toxic agents. The ability to follow changes in the expression of all proteins in a tissue or cell type rather than one, two or perhaps several as is commonly undertaken at present will dramatically increase our understanding of each of these processes. Similarly, analysis of the differences in the proteomes of pathogenic versus nonpathogenic bacterial strains should prove valuable. Of course, proteome analysis will also be used to characterize all the posttranslational modifications, including phosphorylation, fatty acylation, acetylation, ADP-ribosylation and glycosylation, that occur and can change in each of the circumstances detailed above.

It is becoming increasingly apparent that the transmission of information within cells and many other diverse subcellular activities (vesicle transport, transcription, mRNA processing, transport across the nuclear membrane) involve complex and dynamically changing protein assemblies. Proteome analysis will assist in identifying the components of complex protein assemblies either by direct analysis of the protein complexes or, for example, by analysis of immunoprecipitates of one of the protein components of a complex. By revealing which proteins are expressed within a particular cell, it will also assist in interpreting whether potential combinations of protein interactions identified by a two-hybrid screen actually occur in individual cells. In this way, 2-DE and proteome analysis should play a vital role in the generation of cell-specific protein linkage maps (see Ref. 27). Furthermore, with appropriate prefractionation of samples, it will also be possible to analyse the subcellular localization and potential changes in subcellular localization of proteins. This will be particularly useful for comprehensive analysis of changes in the localization of intracellular signalling proteins (see Refs 28 and 29). Many of the potential applications of proteome analysis were discussed

recently at a meeting ('From Genome to Proteome') held in Siena, Italy (16–18 September 1996) and are reported in a special issue of *Electrophoresis* published in April 1997.

Conclusions and future prospects

It is evident that recent years have seen such dramatic and exciting advances in 2-DE and methods to characterize gel-separated proteins that the aim of screening comprehensively the proteins expressed in an organism, a tissue or particular cell type is now becoming an achievable goal. However, a number of technical challenges will have to be overcome before proteome analysis realizes its full potential. Importantly, in the majority of current applications, the 2-D gels are revealing only proteins of high-to-moderate abundance, and technical improvements will be needed to increase the sensitivity of protein detection. The recent surge of interest and activity in the field of proteome analysis greatly increases the likelihood that these developments will be made.

In the same way that the increase in DNA sequence data raised a number f important issues, so too will the future expansion in protein data. These issues include ensuring the accuracy of pr tein characterizations and the validity of reported changes in protein expression; safeguarding access to data; and collaborative management and integration of projects to avoid unnecessary duplication. Much has been learnt from the genome projects, and we hope that this article will enhance awareness of proteome-analysis projects and assist in the formation of a proteome-analysis network to plan, manage and integrate existing and future proteome projects.

References

1 BULT, C. J. *et al.* (1996) *Science* 273, 1058–1073
2 VENTER, J. C., SMITH, H. O. and HOOD, L. (1996) *Nature* 381, 364–366
3 JOHNSTON, M. (1996) *Curr. Biol.* 6, 500–503
4 OLIVER, S. (1996) *Nature* 379, 597–600
5 MANN, M. and WILM, M. (1995) *Trends Biochem. Sci.* 20, 219–224
6 PATTERSON, S. D. (1994) *Anal. Biochem.* 221, 1–15
7 PATTERSON, S. D. and AEBERSOLD, R. (1995) *Electrophoresis* 16, 1791–1814
8 JUNGBLUT, P. *et al.* (1996) *Electrophoresis* 17, 839–847
9 SCHENA, M., SHALON, D., DAVIS, R. W. and BROWN, P. O. (1995) *Science* 270, 467–470
10 VELCULESCU, V. E., ZHANG, L., VOGELSTEIN, B. and KINZLER, K. W. (1995) *Science* 270, 484–487
11 CHEE, M. *et al.* (1996) *Science* 274, 610–614
12 SMITH, V., BOTSTEIN, D. and BROWN, P. O. (1995) *Proc. Natl. Acad. Sci. U. S. A.* 92, 6479–6483
13 CELIS, J. E. *et al.* (1995) *Electrophoresis* 16, 2177–2240
14 ANDERSON, N. L., TAYLOR, J., HOFMANN, J-P., ESQUER-BLASCO, R., SWIFT, S. and ANDERSON, N. G. (1996) *Toxicol. Pathol.* 24, 72–76
15 WILKINS, M. R. *et al.* (1995) *Biotechnol. Genet. Eng. Rev.* 13, 19–50
16 BJELLQVIST, B. *et al.* (1993) *Electrophoresis* 14, 1357–1365
17 BLOMBERG, A. *et al.* (1995) *Electrophoresis* 16, 1935–1945
18 CORBETT, J. M., DUNN, M. J., POSCH, A. and GORG, A. (1994) *Electrophoresis* 15, 1205–1211
19 ANDERSON, N. G. and ANDERSON, N. L. (1996) *Electrophoresis* 17, 443–453
20 CORDWELL, S. *et al.* (1995) *Electrophoresis* 16, 438–443
21 WHEELER, C. H. *et al.* (1996) *Electrophoresis* 17, 580–587
22 WILKINS, M. R. *et al.* (1996) *Biotechnology* 14, 61–65
23 WILKINS, M. R. *et al.* (1996) *Biochem. Biophys. Res. Commun.* 221, 609–613
24 MANN, M. and WILM, M. (1994) *Anal. Chem.* 66, 4390–4399
25 PATTERSON, S. D., THOMAS, D. and BRADSHAW, R. A. (1996) *Electrophoresis* 17, 877–891
26 WILM, M. *et al.* (1996) *Nature* 379, 466–469
27 EVANGELISTA, C., LOCKSHON, D. and FIELDS, S. (1996) *Trends Cell Biol.* 6, 196–199
28 HUNTER, T. (1996) *Trends Cell Biol.* 6, 254–255
29 VANDROMME, M., GAUTHIER-ROUVIÈRE, C., LAMB, N. and FERNANDEZ, A. (1996) *Trends Biochem. Sci.* 26, 59–64
30 ECKERSKORN, C., JUNGBLUT, P., MEWES, W., KLOSE, J. and LOTTSPEICH, F. (1988) *Electrophoresis* 9, 830–838
31 WASINGER, V. C. *et al.* (1995) *Electrophoresis* 16, 1090–1094
32 PAPPIN, D. J. C., HOJRUP, P. and BLEASBY, A. J. (1993) *Curr. Biol.* 3, 327–332
33 COTTRELL, J. S. (1994) *Peptide Res.* 7, 115–124
34 MATSUDAIRA, P. (1987) *J. Biol. Chem.* 262, 10035–10038
35 ROSENFELD, J., CAPDEVIELLE, J., GUILLEMOT, J. C. and FERRARA, P. (1992) *Anal. Biochem.* 203, 173–179

Acknowledgements

We are grateful to the many colleagues who provided material for the preparation of this article and apologize to those whose work we have been unable to include because of space restraints. S. R. P. acknowledges support from the Wellcome Trust and MRC; M. R. W. acknowledges support from the Helmut Horton Foundation; and M. J. D. acknowledges support from the British Heart Foundation.

BOOK REVIEWS

Branching out in neurobiology

Nerve Growth and Guidance (Frontiers in Neurobiology)

edited by C. D. McCaig, Portland Press, 1996. £60.00 (174 pages) ISBN 1 855 78 085 2

This book is a concise review f r those interested in recent advances in the field of neurobiology and makes excellent supplementary reading for senior college and graduate students.

The integrity of the nervous system depends upon highly specific connections formed between neurons during development and regeneration. The specificity of these connections requires axonal outgrowth along defined pathways. Since the early morphological studies of Ramon y Cajal and Harrison, the control of events such as axonal guidance and target recognition have been attributed to the growth cone, a specialized structure located at the distal tip of growing neurites (Chapter 1).

Growth cones are capable of responding to various environmental signals. Signals such as soluble trophic or tropic fact rs and other molecular cues residing in the extracellular matrix and cell membranes have long been shown to be involved in neurite outgrowth and guidance (Chapters 1, 6 and 8). Snow *et al.* remind us of another potential player, the proteoglycans, that has been increasingly implicated in the development and maintenance of the nervous system (Chapter 6). One remaining puzzle, however, is the question of how the growth-promotion molecules – dozens of them identified so far – are able to govern the thousands of specific connections made among different neurons. Recently, a variety of 'collapsing' factors were added to this fascinating repertoire f regulators, giving a Yin–Yang balance to what we know about axonal navigation and guidance. Bandlow in Chapter 7 and Shewan and Cohen in Chapter 8 summarize